



OPEN ACCESS

EDITED BY

Yuan Lu,
Tsinghua University, China

REVIEWED BY

Yixin Huo,
Beijing Institute of Technology, China
Amir Pandi,
Université Paris Cité, France
Feng Xu,
Xi'an Jiaotong University, China

*CORRESPONDENCE

Hao Qi,
✉ haoq@tju.edu.cn

†These authors have contributed equally
to this work

RECEIVED 06 April 2023

ACCEPTED 17 May 2023

PUBLISHED 25 May 2023

CITATION

Li J, Li P, Liu Q, Li J and Qi H (2023),
Translation initiation consistency
between *in vivo* and *in vitro* bacterial
protein expression systems.
Front. Bioeng. Biotechnol. 11:1201580.
doi: 10.3389/fbioe.2023.1201580

COPYRIGHT

© 2023 Li, Li, Liu, Li and Qi. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Translation initiation consistency between *in vivo* and *in vitro* bacterial protein expression systems

Jiaojiao Li^{1,2†}, Peixian Li^{1,2†}, Qian Liu^{1,2}, Jinjin Li^{1,2} and Hao Qi^{1,2,3*}

¹School of Chemical Engineering and Technology, Tianjin University, Tianjin, China, ²Frontier Science Center for Synthetic Biology (Ministry of Education), Tianjin University, Tianjin, China, ³Zhejiang Shaoxing Research Institute of Tianjin University, Shaoxing, China

Strict on-demand control of protein synthesis is a crucial aspect of synthetic biology. The 5'-terminal untranslated region (5'-UTR) is an essential bacterial genetic element that can be designed for the regulation of translation initiation. However, there is insufficient systematical data on the consistency of 5'-UTR function among various bacterial cells and *in vitro* protein synthesis systems, which is crucial for the standardization and modularization of genetic elements in synthetic biology. Here, more than 400 expression cassettes comprising the GFP gene under the regulation of various 5'-UTRs were systematically characterized to evaluate the protein translation consistency in the two popular *Escherichia coli* strains of JM109 and BL21, as well as an *in vitro* protein expression system based on cell lysate. In contrast to the very strong correlation between the two cellular systems, the consistency between *in vivo* and *in vitro* protein translation was lost, whereby both *in vivo* and *in vitro* translation evidently deviated from the estimation of the standard statistical thermodynamic model. Finally, we found that the absence of nucleotide C and complex secondary structure in the 5'-UTR significantly improve the efficiency of protein translation, both *in vitro* and *in vivo*.

KEYWORDS

cell-free system, gene expression regulation, protein engineering, synthetic biology, *E. coli*

1 Introduction

Tunable genetic elements are crucial for various applications in metabolic engineering and synthetic biology (Chappell et al., 2013). Protein abundance in bacterial cells is determined by factors such as regulatory elements in the RNA, ribosome density (Arava et al., 2005), and selection of the start codon (Basu et al., 2022). Studies have demonstrated that the 5'-untranslated regions (5'-UTRs) (Ding et al., 2018) comprising the core ribosome-binding site (RBS) is a pivotal element controlling the translation initiation efficiency and protein expression level (Salis et al., 2009; Shi et al., 2018) (Seo et al., 2013). Recent studies found that designed 5'-UTRs can control protein translation at a wide range of levels in various expression systems, including yeast (Cuperus et al., 2017), *E. coli* (Duan et al., 2022), and human cells (Sample et al., 2019). Interestingly, these studies indicated that UTRs with a biased nucleotide composition are more conducive to high translation efficiency. In particular, A and U in the UTR sequences significantly contribute to the translation efficiency.

E. coli cell is the most common model system for the efficient expression of heterologous proteins because of its well-studied genetic background, fast growth rate, simple genetic manipulation, and high protein expression capacity (Yoon et al., 2012; Marisch et al., 2013; Han et al., 2020). The suitability of *E. coli* for high cell-density culture enables the cheap production of recombinant proteins (Phue et al., 2005; Son et al., 2011; Englaender et al., 2017), while the well-studied metabolic network makes *E. coli* a popular platform for natural product synthesis (Pontrelli et al., 2018). The K-12 and B strains of *E. coli* with their variants are among the most widely used host cells in biological research and industrial fermentation (Marisch et al., 2013), whereby *E. coli* K-12 strains are more widely used in genetic manipulation and biochemical research (Kuhnert et al., 1995; Posfai et al., 2006). According to Northern blot analysis, the mRNA expression level of the *E. coli* B strain is higher than in the K-12 strain (Shiloach et al., 2000; Phue and Shiloach, 2004). Furthermore, *E. coli* B strains, which lack genes for flagella production and express the motility-related proteins at a relatively low level, can achieve a higher biomass yield and faster growth than *E. coli* K-12 strains (Yu et al., 2002; Shiloach et al., 2010). Previous studies have investigated the differences between *E. coli* B and K-12 strains by transcriptomic and proteomic approaches (Marisch et al., 2013), as well as metabolomic analysis (Shiloach et al., 2010).

In vitro protein synthesis, CFPS provides high flexibility and controllability. Cell extract-based *in-vitro* systems are used as a platform for rapid characterization of regulatory elements, circumventing the inherently time-consuming and low-throughput manipulation of living cells, in various applications of proteins engineering and synthetic biology (Hodgman and Jewett, 2012; Sun et al., 2014; Lu, 2017). In contrast to the complexity and limitations of living cells, the functional performance of genetic elements can be rapidly validated and screened from libraries using *in vitro* platforms, such as PURE (Shimizu et al., 2001) and lysate-based systems (Kightlinger et al., 2018). Extract-based systems have been utilized to rapidly evaluate different RBSs, showing the inconsistency between the theoretical prediction and actual protein synthesis levels (Wang et al., 2018). Interestingly, recent studies showed that there was also a deviation between the *in vivo* and *in vitro* protein synthesis levels using the same elements (Zhang et al., 2021). Therefore, the systematic characterization of the regulation of protein expression in various strains and even *in vitro* systems is necessary for improving standardization and modularization in synthetic biology.

In this study, we systematically quantified the expression levels of 416 superfolder GFP (sfGFP) expression cassettes with various 5'-UTR sequences in two *E. coli* strains, BL21 Star (DE3) and JM109 (DE3), as well as an *in vitro* lysate-based system to evaluate the consistency of UTR function. In contrast to the strong correlation between these two cellular systems, the consistency between *in vivo* and *in vitro* protein translation is lost, whereby both *in vivo* and *in vitro* translation levels evidently deviate from the estimation of the standard statistical thermodynamic model. Specifically, we also quantified the function of 5'-UTRs comprising only three types of nucleotides. Finally, we found that the absence of C nucleotides and a complex secondary structure in the 5'-UTR significantly improve protein translation, both *in vivo* and *in vitro*. Therefore,

we believe that designing genetic elements with high consistency across various strains and even *in vitro* systems is necessary for effective standardization and modularization in synthetic biology.

2 Materials and methods

2.1 Bacterial strains, plasmids, and growth conditions

The *E. coli* strain JM109 (DE3) [*endA1*, *glnV44*, *the-1*, *relA1*, *gyrA96*, *recA1*, *mcrB+*, Δ (*lac-proAB*), *e14-* (*F'* *traD36 proAB*, *lacIqZΔM15*), *hsdR17* (*rk⁻*, *mk+*), + λ (DE3)] and the *E. coli* strain BL21 Star (DE3) [*F⁻ompT hsdS_B* (*r_B⁻*, *m_B⁻*) *galdcmrne131* (DE3)] were used as host organisms and cultivated in Luria–Bertani (LB) medium supplemented with the working concentration of 100 μ g/mL ampicillin at 37°C for the 5'-UTRs activity validation. Primers for amplification were synthesized by Azena. The pUC19 plasmid and pUC19-wt-sfGFP plasmid were constructed by our laboratory, and the recombinant plasmids with 5'-UTR sequences were verified by Sanger sequencing. All primers are listed in Supplementary Table S1.

2.2 Construction of randomized libraries

We generated a library of plasmids with 5'-UTRs containing a 25 nucleotides long randomized region directly upstream of the start codon. Firstly, the upstream primer 25N-F and downstream primer 25-R were synthesized, annealed and amplified using the Primer STAR Max premix (TaKaRa). The PCR was performed with one cycle of 98°C (10 s), 55°C (5 s), and 72°C (1 h) followed by cooling to 10°C. Secondly, the linearized vector was obtained from the original plasmid pUC19-nonRBS-sfGFP constructed in our laboratory by PCR amplification using the V19-F and V19-R. Then, the amplified PCR products were purified using a Tian quick PCR Purification Kit (Qiagen) or Tian quick Gel Extraction Kit (Qiagen). Finally, the recombinant plasmids were assembled using the ClonExpress II One Step Cloning Kit (Vazyme Biotech Co., Ltd.) and transferred into the competent *E. coli* JM109 (DE3) cells. Libraries were cultivated overnight on the LB plate with ampicillin for until single colonies appeared. The cultured plates were placed in a refrigerator at 4°C for the subsequent activity measurements.

2.3 Characterization of 5'-UTRs in live *E. coli* cells

The growth and fluorescence measurements were performed in 96-well high-throughput format. Single colonies were picked from the plate and cultivated overnight in LB medium with 100 μ g/mL ampicillin at 37°C, 14,000 rpm for 12 h in the deep well maximizer (TAITEC, MBR-022UP). Then, 160 variants with different 5'-UTR regions were selected by Sanger sequencing to remove the repeat sequences. There were three positions (wells) in each 96-deep-well plate for the controls, including the negative control in the H10 position, the positive control in the H11 position and the blank control in the H12 position. The cells were cultured overnight,

and seeded at a ratio of 1:100 into the 96 deep-well plate (VIOX scientific) containing 300 μL liquid LB medium with 100 $\mu\text{g}/\text{mL}$ ampicillin and 0.5 mM IPTG per well, followed by culture at 37°C, 14,000 rpm for 5 h. After that, 80 μL of culture was taken out with an electronic pipette (Eppendorf Xplorer[®]) and placed it into a 96-well plate (Corning#3762) which contained 120 μL of LB medium in each well. We measured the optical density (OD_{600}) and fluorescence value of each culture (485/535 nm) using a microplate reader (Tecan Spark multimode microplate reader). In the process of activity measurement, we measured each sample three times and calculated its average value. For the correlation analysis of 5'-UTRs in the living cell system, the absolute fluorescence for each 5'-UTR sequence was calculated and normalized to the OD_{600} . For the correlation analysis of 5'-UTRs between *in vivo* and *in vitro* systems, we respectively calculated their relative fluorescence according to the formula (Salis, 2011):

$$\text{FLU}_{\text{sample}} = \frac{\text{FLU}_{\text{sample}} - \text{FLU}_{\text{media}}}{\text{OD}_{\text{sample}} - \text{OD}_{\text{media}}} - \frac{\text{FLU}_{\text{pUC19}} - \text{FLU}_{\text{media}}}{\text{OD}_{\text{pUC19}} - \text{OD}_{\text{media}}} \quad (1)$$

where the $\text{FLU}_{\text{sample}}$, $\text{FLU}_{\text{media}}$, and $\text{FLU}_{\text{pUC19}}$ respectively represent the sfGFP fluorescence of the sample, the blank control, and the negative control, while the $\text{OD}_{\text{sample}}$, OD_{media} , and OD_{pUC19} respectively represent the OD_{600} of the sample, the blank control, and the negative control.

The relative fluorescence value per cell concentration of wild-type cells was calculated according to the formula:

$$\text{FLU}_{\text{wt}} = \frac{\text{FLU}_{\text{wt}} - \text{FLU}_{\text{media}}}{\text{OD}_{\text{wt}} - \text{OD}_{\text{media}}} - \frac{\text{FLU}_{\text{pUC19}} - \text{FLU}_{\text{media}}}{\text{OD}_{\text{pUC19}} - \text{OD}_{\text{media}}} \quad (2)$$

where FLU_{wt} represents the sfGFP fluorescence of the positive control, and OD_{wt} represents the OD_{600} of the positive control.

The relative activity of each sample was standardized to the FPLC_{wt} for the comparisons of different 5'-UTRs. Here, we defined the relative intensity of the 5'-UTR region as the percentage of the relative fluorescence value per cell concentration of the sample and the relative fluorescence value per cell concentration of wild-type cells, according to the formula:

$$\text{P\%} = \frac{\text{FPLC}_{\text{sample}}}{\text{FPLC}_{\text{wt}}} \times 100\% \quad (3)$$

According to the above calculation method, we measured and analyzed the real relative activity of 93 sequences with different 5'-UTRs in the *E. coli* JM109 (DE3) and BL21 Star (DE3) strains.

2.4 Batch extraction of plasmids and bacterial transformation

The direct boiling method for plasmid DNA extraction was adopted as described before (Peng et al., 2013), with minor modifications as follows. The cryostock containing the strain library was re-cultured at 37°C and 14,000 rpm for 12 h in the 96 deep-well plate. Then, 70 μL samples from the 96 deep-well plate were added into eight consecutive rows of PCR tubes with corresponding position labels. The *E. coli* cells were collected by centrifugation at 6,000 rpm for 10 min with a high-speed refrigerated microcentrifuge (MDX-310, LTD.), and added to 100 μL of ddH₂O, followed by vigorous vortexing to

homogenize the suspension. Then, the cell suspensions were incubated at 95°C for 10 min and subsequently centrifuged at 14,000 rpm at room temperature for 15 min. Finally, 50 μL of the supernatant were transferred to a clean 96-well PCR plate and stored at -40°C.

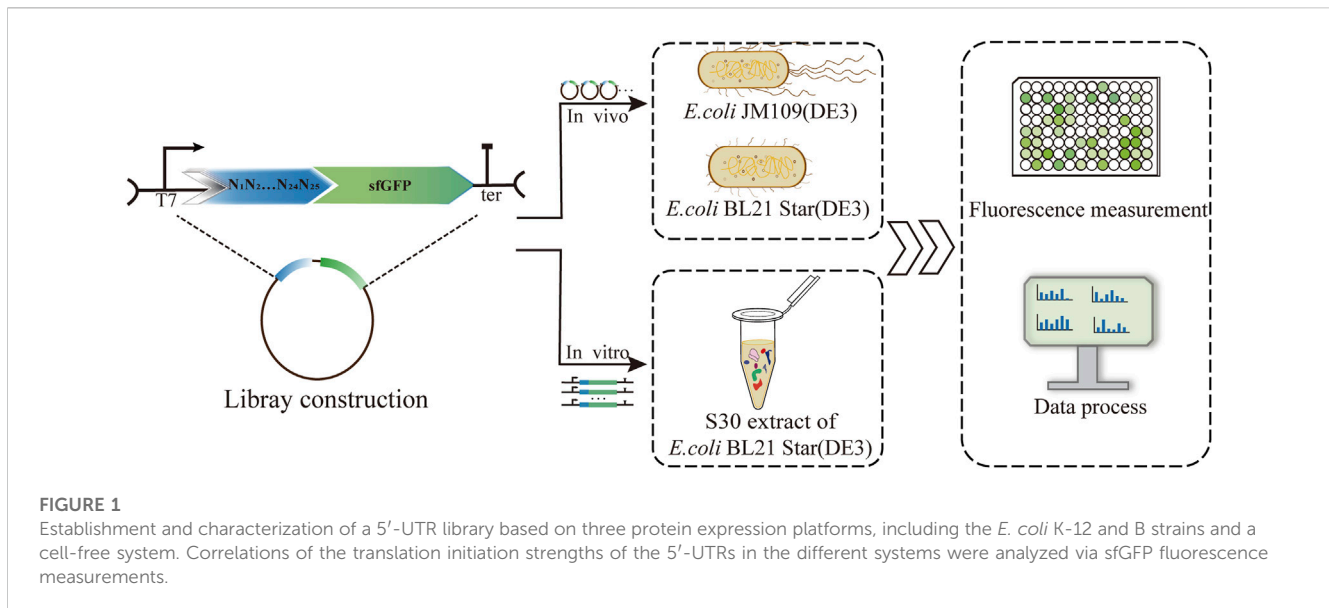
The competent *E. coli* BL21 Star (DE3) cells were prepared by the calcium chloride method (Seidman et al., 2001), and packed into 96-well plates at 100 $\mu\text{L}/\text{well}$ for transformation. The amount of the obtained plasmid plays an important role in the successful transfer process, considering the purity and the transformation efficiency. Hence, we carried out a series of optimization experiments and determined the optimal amount (Supplementary Figure S1A). Then, the competent cells were incubated on ice for 30 min. Heat-shocked for 45 s at 42°C in an electric constant temperature water bath (DK-98-IIA Tianjin TEST Instrument Co., Ltd.), and immediately placed in an ice bath for 2 min. Subsequently, 700 μL SOB medium was added and the cells were recovered at 37°C and 14,000 rpm for 3 h. Finally, 30 μL of the cells were pipetted into 300 μL of liquid LB medium containing 100 mg/mL ampicillin and cultured overnight at 37°C and 14,000 rpm.

2.5 Characterization of 5'-UTRs in a cell-free system

The S30 cell extract was prepared as described in our previous study (Wu et al., 2022), but we did not carry out the run-off reaction for high CFPS yields. Hence, the subpackage extract was directly frozen in liquid nitrogen and stored in the refrigerator at -80°C.

To prepare the DNA template for CFPS reaction, a standard PCR reaction for each sample in a 96-well plate was performed in a 25 μL system comprising 1 \times Easy Taq Buffer (TransGen, ET101), 0.2 mM dNTPs, 2.5 U Easy Taq polymerase (TransGen Biotech), 1 μL extracted DNA template, 0.2 mM forward primer F2 and 0.2 mM reverse primer R2. Thermocycling conditions were as follows: 10 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 55°C, and 60 s at 72°C, followed by a 5 min extension at 72°C. The PCR products were analyzed by electrophoresis on a 2% agarose gel to confirm the size of the band and subsequently stored in the refrigerator at 4°C.

For the characterization of 5'-UTR *in vitro*, a standard CFPS reaction in the 20 μL reaction system was used containing 2 μL of PCR products and other reaction components same as described in a previous report (Wu et al., 2022). The addition of 2 μL of the PCR products was optimal for achieving the highest yield in this CFPS reaction (Supplementary Figure S4B). All reactions were correspondingly pipetted into a 384-well plate (Corning #3762) and transferred to a microplate reader with real-time fluorescence monitoring (excitation at 485 nm and emission at 535 nm) for up to 3 h in 5-min intervals at 30°C, shaking linearly for 5 s before each measurement. The reported data are the averages of three independent measurements. The background fluorescence intensity from a cell-free reaction with the empty plasmid (pUC19) was subtracted from each sample fluorescence measurement, and the resulting intensity values were normalized to the positive control plasmid (pUC19-wt-GFP) to calculate the relative expression strength.



2.6 Analysis of 5'-UTRs sequences

The sequence logo for the 5'-UTR library was generated using the online analysis tool WebLogo (<http://weblogo.berkeley.edu/>). The translation initiation strength of each 5'-UTR was predicted using the RBS Calculator (Salis et al., 2009). For the secondary structure analysis of 5'-UTRs, minimal folding energies (MFE) were calculated for the region encompassing the entire 25 nt 5'-UTR and 27 nt downstream of the sfGFP coding region using NUPACK 4.0. To estimate the recognition of the 5'-UTR region by the 16S rRNA 3'-terminal region, we also calculated the free hybridization energy of 5'-UTR fragments with the anti-SD sequence (5'-ACCUCCUUA-3') using NUPACK 4.0 (Zadeh et al., 2011).

2.7 Statistical analysis

All r values are Pearson correlation coefficients of the strength of the linear relationship between the two sets of data. All scatter plots and histograms were generated using the Origin software package. Quantitative data are presented as means \pm standard deviations (SD) from three experiments.

3 Results and discussion

3.1 Characterization of 5'-UTRs in different bacterial strains

For systematic quantification of the consistency of translation efficiency of 5'-UTRs, we established different protein expression platforms, including the living cell system and cell-free system. The 5'-UTRs library was constructed and measured correspondingly the protein initiation expression levels. Correlations of the translation initiation strengths of the 5'-UTRs in the different systems were analyzed via sfGFP fluorescence measurements (Figure 1). For the

validation of 5'-UTRs consistency, the *E. coli* K-12 and B strains were selected as the protein expression platforms *in vivo*. To assess the translation efficiency in *E. coli* K-12 and *E. coli* B, we measured the fluorescence and analyzed their correlation. According to the ribosome profiling studies (Kim et al., 2012), the average length of 5'-UTRs in *E. coli* cells is 25–30 nt. Firstly, we constructed a library of 5'-UTRs, with unbiased 25 N random nucleotides, where N stands for any of A/C/G/T. The 5'-UTRs library was cloned into a plasmid encoding the sfGFP reporter gene to determine the translation initiation strength by measuring the fluorescence. Therefore, the unbiased 5'-UTR library was introduced into *E. coli* JM109 (DE3) cells cultured on solid LB medium with ampicillin. Then, 160 single colonies with unique 5'-UTRs sequences were randomly picked from the plates for the analysis of sfGFP translation efficiency by measuring the corresponding fluorescence.

To analyze the correlation of the translation initiation strength of the 5'-UTRs in different strains, the plasmids containing the 160 5'-UTRs variants were extracted and transferred into *E. coli* BL21 Star (DE3) cells in batch (see Materials and methods 2.4 for a specific description). The optimal volume of the plasmid DNA of the transformation was quantified to be 8 μ L (Supplementary Figure S1A). For each cell sample, the average fluorescence normalized by cell density (measured absorbance at 600 nm) was quantified for the sfGFP translation efficiency. We plotted the distribution of absolute fluorescence of the 160 unbiased 5'-UTRs sequences measured in both *E. coli* JM109 (DE3) and *E. coli* BL21 Star (DE3) cells (Figure 2A). The absolute fluorescence of the two strains overall spanned a similar range. In addition, the number of BL21 Star (DE3) cells with high activity was significantly higher than JM109 (DE3) cells. The analysis of the activity distribution for all 160 unbiased 5'-UTRs sequences revealed that *E. coli* B strains, as chassis cells, were more favorable than K-12 for the production of heterologous proteins. This result could be explained by the inherent properties of the *E. coli* B strain, as was previously indicated through the comprehensive analysis of multi-omics data,

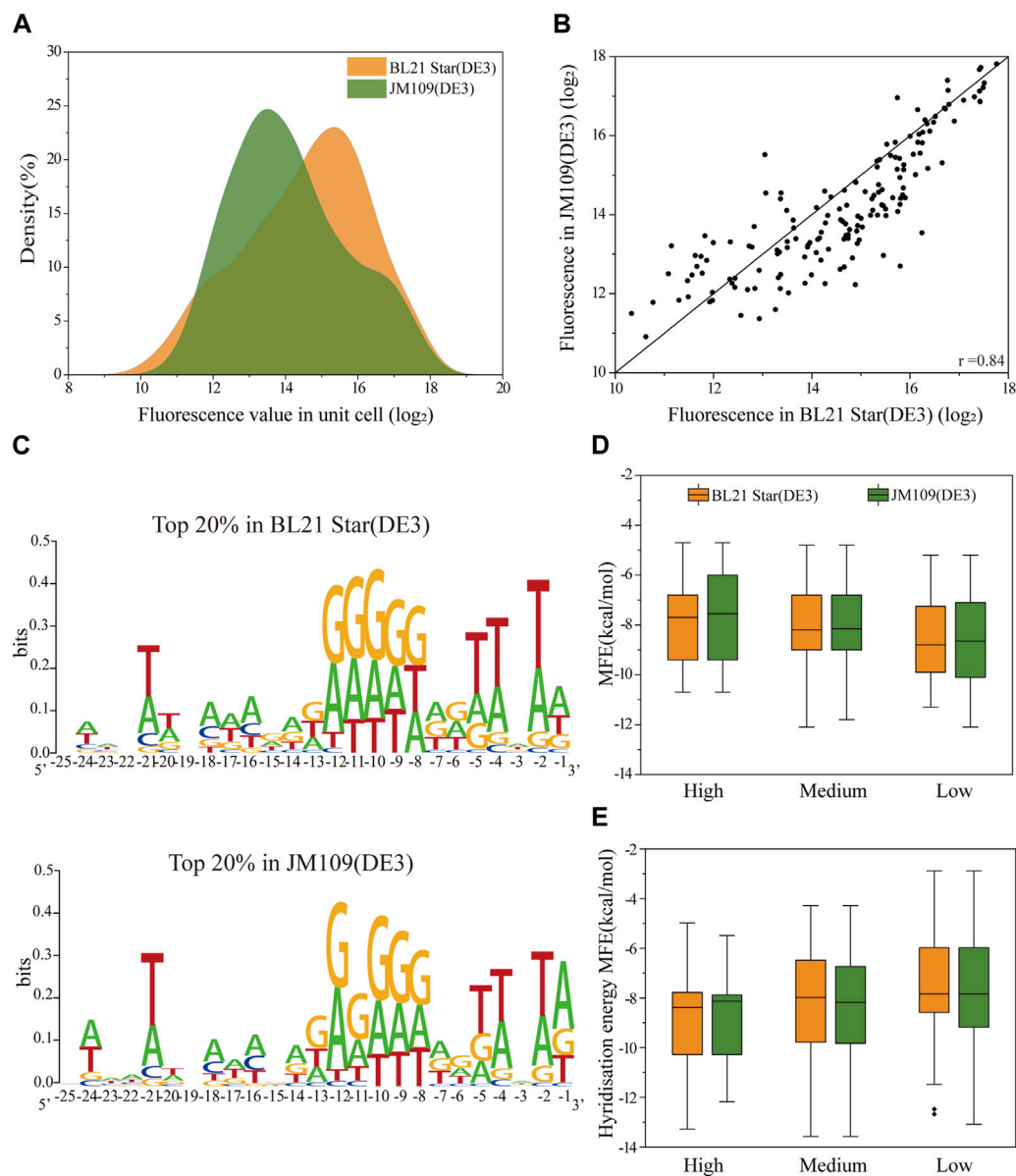


FIGURE 2

The correlation of the translation initiation strengths of the unbiased 5'-UTR library in *E. coli* JM109 (DE3) and BL21 (DE3) *in vivo*. **(A)** The distribution of absolute fluorescence values of the unbiased 5'-UTR library containing 160 unique sequences in both strains. **(B)** The correlation of absolute fluorescence of 160 5'-UTR variants between the two strains (Pearson's $r = 0.84$). **(C)** Sequence logos were calculated for the Top 20% tested sequences in each strain. **(D)** Distributions of the minimal free energy (MFE) of the secondary structure folding for 5'-UTRs with different fluorescence levels between the two strains. **(E)** Distributions of the minimal hybridization energy of the 5'-UTR fragments with the anti-SD sequence ACCUCCUUA at the 3' end of the 16S rRNA between the two strains. We defined absolute fluorescence levels above 16.2 as high activity, those absolute fluorescence levels between 13.2 and 16.2 as medium activity, and those below 13.2 as low activity.

including the genome, transcriptome, proteome, and phenome data (Yoon et al., 2012). The elements involved in the amino acid biosynthesis pathway and the absence of genes for flagella and proteases could contribute to the beneficial metabolism and physiological state of *E. coli* B.

More importantly, with regard to the 5'-UTRs translation regulation, we found a significant correlation in between the *E. coli* JM109 (DE3) and BL21 Star (DE3) cells (Pearson's $r = 0.84$; Figure 2B). It implied that the regulatory properties of the 5'-UTRs in *E. coli* K-12 or B strains are universal. Considering the measured

absolute fluorescence, we measured the respective growth curves of each strain and the strain containing the benchmark plasmid (Supplementary Figure S1B). From the perspective of the strain's inherent characteristics, the growth rates were consistent, but showed that strain B grew a little faster than strain K. (Luli and Strohl, 1990). In contrast, the *E. coli* JM109 (DE3) cells transformed with the positive plasmid grown faster than the corresponding *E. coli* BL21 Star (DE3) cells. In general, the *E. coli* B cells were primarily intended for the synthesis of exogenous recombinant proteins, which is why it has a lower growth rate and tends to use more energy for protein synthesis

(Bentley et al., 1990; Carneiro et al., 2013). The high consistency of the expression levels in the two strains was attributed due to a number of reasons. Firstly, there is more than 92% similarity among aligned regions of the genome between *E. coli* JM109 (DE3) and BL21 Star (DE3) (Jeong et al., 2009). Moreover, similar regulation in *E. coli* K-12 and B strains was previously demonstrated through comprehensive transcriptomic and proteomic data analysis (Marisch et al., 2013). Overall, the translation initiation strength of 5'-UTRs in both industrial hosts showed high similarity, which provides a valuable understanding for designing microbial cell factories in the future.

Previous analyses have demonstrated that mRNA secondary structures as well as the context and position of 5'-UTRs influence the efficiency of protein translation (Evfratov et al., 2017; Duan et al., 2022). To elucidate the high correlation between the two strains, we implemented and analyzed several elements, encompassing context-dependence and secondary structures. We calculated sequence logo for total of 160 5'-UTRs sequences after sequencing the corresponding regions from all the strains (Figure 2C). It is generally understood that adenosine and guanine showed high conservation levels at positions -7 to -12, approximately close to 0.3 bits. As expected, the characteristics of the sequence in this conserved region are consistent with SD sequence features and may be an SD-like sequence (Li et al., 2012). In addition, we also respectively analyzed the nucleotide frequencies at positions -25 to -1 relative to the start codon for the top 20% sequences in the total random library for differences between *E. coli* JM109 (DE3) and *E. coli* BL21 Star (DE3) (Figure 2C). This visible result demonstrated that compositional bias toward A and T at positions -2, -4, and -5 of the 5'-UTR were the consistent characteristic in both strains, which demonstrated that A-U enhancer interacted with ribosomal protein S1 to promote protein expression (Komarova et al., 2005; Duan et al., 2022). As expected, 5'-UTR variants with higher protein expression contained sequences more similar to the SD and SD-like sequences. Overall, the 5'-UTR sequence contents dependence in both strains was generally adapted for the precise tuning of protein expression.

Next, we assessed whether there was a consistent effect of secondary structure on the translation efficiency for protein expression in the K-12 and B strains. To calculate the predicted minimum free energy (MFE) of the 5'-UTRs, we used the NUPACK algorithm to fold the 5'-UTRs sequence along with the first 27 nt of the sfGFP coding region. We classified the protein expression levels into three sets of high, median, and low expression, and then plotted the relationship between MFE and the protein expression level. Binning the 5'-UTRs by their MFE, we found that the higher MFE fraction corresponded to increased protein expression in both K-12 and B strains, implying no difference in the effect of MFE on translation in the two different strains (Figure 2D). Hence, a stable secondary structure of mRNA might mostly downregulate the protein expression, which was not affected by the properties of the cells themselves. In addition, the binding interaction between the SD region of 5'-UTR and the 3' end regions of 16S rRNA could primarily determine the translation initiation process (Shine and Dalgarno, 1974). The free hybridization energy of the 5'-UTR fragments with the anti-SD sequence ACCUCCUUA at the 3' end of the 16S rRNA was calculated for the correlation analysis of protein expression between the *E. coli* JM109 (DE3) and BL21 Star (DE3) strains. Lower hybridization energy was correlated with

higher fluorescence in both *E. coli* strains (Figure 2E). These results demonstrated that the interaction of the ribosomes with the 5'-UTRs did not result in any difference in the evaluation of protein expression between *E. coli* K-12 and B. In addition to the local mRNA structure, the N-terminus sequence around the translation start sites in coding protein regions affected the translation initiation efficiency (Kudla et al., 2009; Goodman et al., 2013). Hence, we selected five 5'-UTR sequences with high, medium and low translation efficiency respectively and fused them with another protein, Glutathione-S-transferase (GST), to evaluate the translation initiation consistency in *E. coli* BL21 Star (DE3) strains. In *E. coli* B cells, the expression of GST and sfGFP showed a medium correlation (Pearson's $r = 0.75$; Supplementary Figure S2). This result suggested that the regulation consistency of the same 5'-UTR sequence's translation initiation efficiency for different reporter proteins was correlated, but the correlation consistency was lower than that for different types of cells under the same reporter protein condition. This indicated that considered codon distribution in different reporter protein sequences may play a crucial role in the consistency of 5'-UTR regulation (Verma et al., 2019). In summary, the high correlation of the 5'-UTR characteristics between the K-12 and B strains in multiple data analyses provided a general rule for the assignment of 5'-UTR elements.

3.2 The nucleotide content of the 5'-UTR influences translation initiation

To decipher the correlations among biased 5'-UTR library characterization in the two *E. coli* strains, four biased 5'-UTR libraries were designed by introducing 25 degenerate bases (B/H/V/D), where C/G/T was designated as B, A/C/T was designated as H, A/C/G was designated as V, and A/G/T was designated as D (Figure 3). The four resulting biased 5'-UTRs libraries were first introduced into the JM109 (DE3) cells, and picked colonies were then sequenced and then transferred into the BL21 Star (DE3) strain (see methods) as well. Finally, the absolute fluorescence of the four biased 5'-UTR libraries in both strains was measured for the correlation analysis. Notably, we discovered that there was a high variability in the consistency of the translation initiation strength between the two strains for these four libraries (Figure 3). Especially the 25H library showed a lower association with the protein expression in both *E. coli* strains (Pearson's $r = 0.38$ for 25H library), and the corresponding fluorescence levels were relatively low, whereby the JM109 (DE3) strains spanned a narrow range level (Supplementary Figure S3B). Moreover, the weak performance of the 25H library illustrates that guanine dependence of 5'-UTRs can lead to remarkable variations of protein translation in different chassis hosts. There is an urgent need for 5'-UTRs with low expression for the engineering of the metabolic pathways, which can be rapidly selected from a biased library, such as the above-mentioned 25B library. However, the 25B and 25V libraries showed a higher correlation of protein expression in both *E. coli* strains (Pearson's $r = 0.84$ for 25B, Pearson's $r = 0.88$ for 25V). More importantly, the 25B and 25V libraries showed the interval distribution of measured activity in two chassis strains was uniform, which is close to the normal distribution

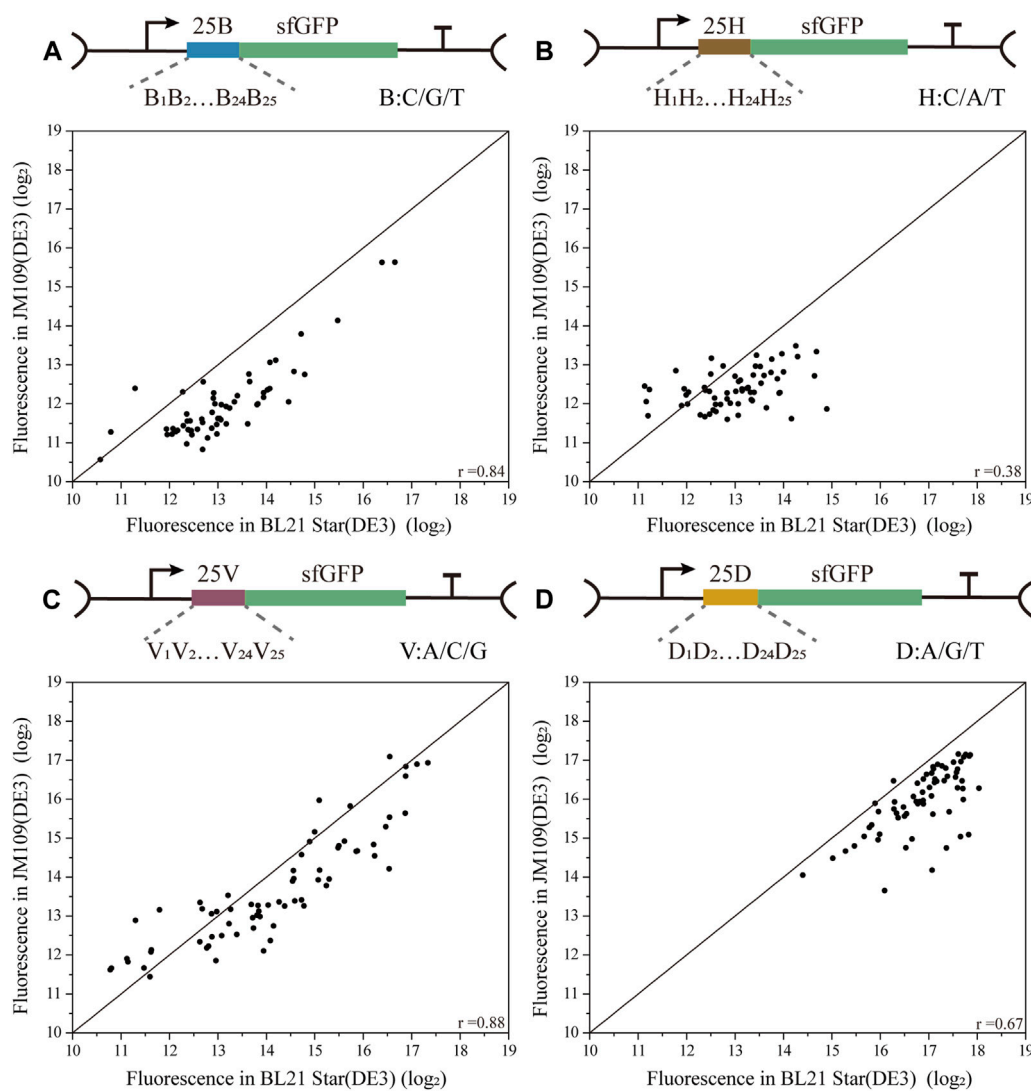


FIGURE 3

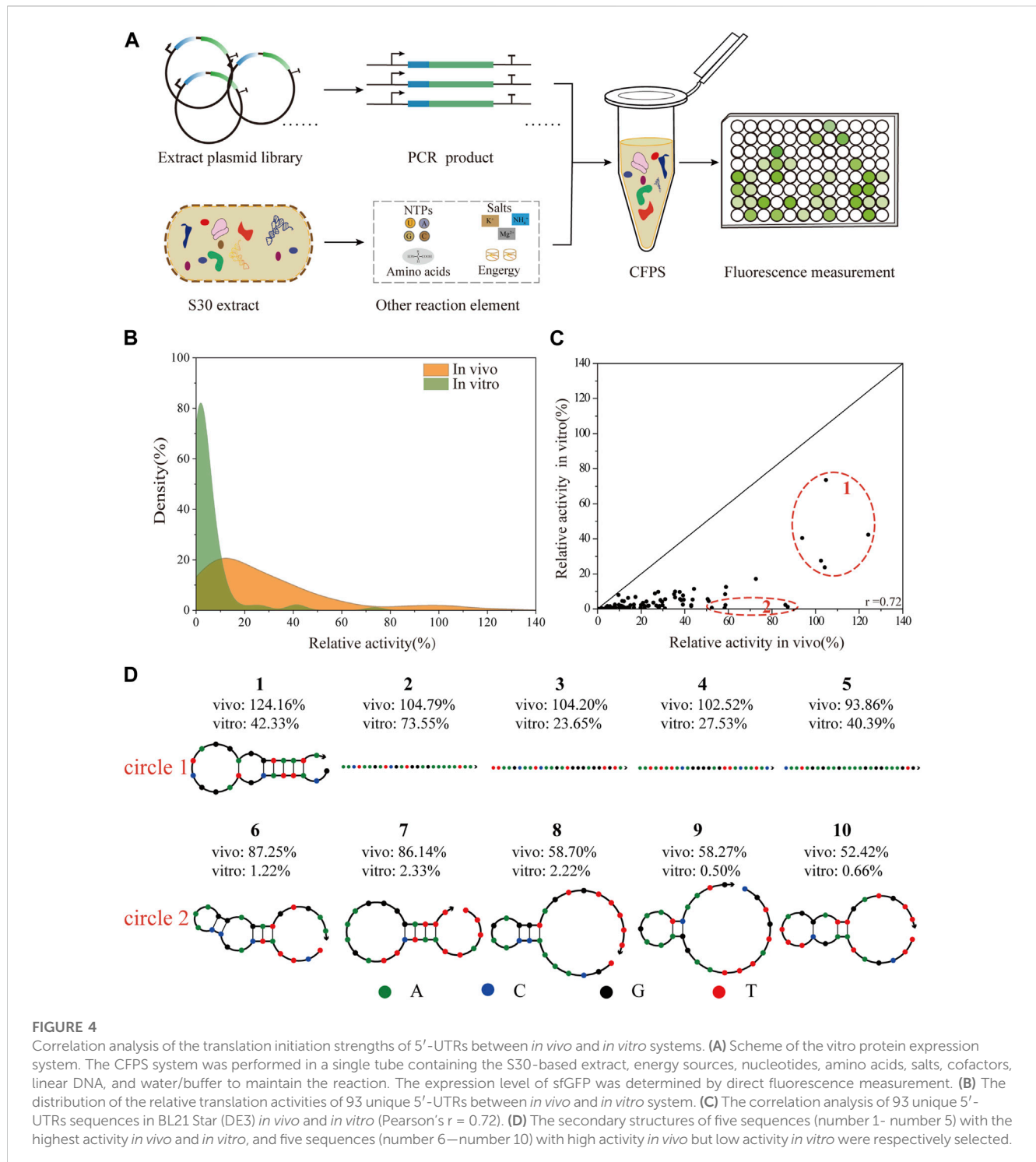
Correlation analysis of four biased 5'-UTR libraries in *E. coli* BL21 Star (DE3) and JM109 (DE3). (A) The 25 B library contained 59 unique 5'-UTRs with 25 continuous degenerate bases B, where B stands for C/G/T (Pearson's $r = 0.84$). (B) The 25H library contained 63 unique 5'-UTRs with 25 continuous degenerate bases H, where H stands for C/A/T (Pearson's $r = 0.38$). (C) The 25V library contained 68 unique 5'-UTRs with 25 continuous degenerate bases V, where V stands for A/C/G (Pearson's $r = 0.88$). (D) The 25D library contained 66 unique 5'-UTRs with 25 continuous degenerate bases D, where D stands for A/G/T (Pearson's $r = 0.67$). The correlation analysis of these four libraries between BL21 Star (DE3) and JM109 (DE3) is based on the analysis and comparison of absolute fluorescence values.

(Supplementary Figure S3A, S3C). In other words, the 5'-UTR sequences lacking A and T exhibited high versatility between *E. coli* K-12 and B strains for the measurement of translation efficiency. This interesting phenomenon means that A and T context dependence contributes weakly to the correlation between the biased library and host cells. Significantly, the 25D library of 5'-UTRs performed with medium consistency in protein expression levels between K-12 and B strains (Pearson's $r = 0.67$ for 25D), which is shown that mainly exhibit high fluorescence levels (Supplementary Figure S3D). In agreement with a previous study (Evfratov et al., 2017), a low proportion of cytidine residues promoted the efficiency of translation. This result reinforces the notion that target sequences with higher translation efficiency might be screened rapidly from a biased library of 5'-UTRs with the

deletion of cytidine residues, which reduced the time and cost of optimization to a certain extent. Therefore, we demonstrated that there is an obvious bias in the distribution and consistency of the regulatory strength of the 5'-UTR sequences for translation in different cells.

3.3 Comparison of 5'-UTR functions between *in vivo* and *in vitro* protein synthesis

To facilitate the engineering of biological elements for cell-free systems, we tested the consistency of 5'-UTRs regulation *in vivo* and *in vitro* (Figure 4A). With the goal of constructing a robust, high-yielding CFPS system, a reported protocol (Wu et al., 2022), was



adopted to prepare S30-extracts derived from the cells of *E. coli* JM109 (DE3) and BL21 Star (DE3). To assess the potential ability of protein synthesis, we subsequently carried out the CFPS of the standard sfGFP template in a 20 μ L batch reaction for 3 h at 30°C. Representative time courses of sfGFP synthesis using the purified PCR products with online fluorescence measurements were shown in [Supplementary Figure S4A](#). According to the experimental data, the protein synthesis level of the BL21 Star (DE3) extract system was

almost 10-fold higher than that of the JM109 (DE3) lysate system. Consistent with previous studies ([Kwon and Jewett, 2015](#)), the extract from the *E. coli* K-12 strain showed a lower protein expression level, which could be improved by modifying a few parameters. The S30 extract derived from the BL21 Star (DE3) cells could inherently enable higher protein yields, owing to its genome modification ([Jiang et al., 2021](#)). Therefore, we applied the *E. coli* BL21 Star (DE3) extract-based CFPS system to validate the

correlation of the whole library of 93 5'-UTR variants between the living cell system and *in vitro* system. After demonstrating high-yield protein expression, we set out to test the feasibility of directly using the PCR amplicon as the CFPS template for high-throughput synthesis. In the gradient addition of the PCR amplicons, we found that 2 μ L of the PCR products resulted in close to 49% of the sfGFP synthesis level obtained using the purified PCR product (Supplementary Figure S4B). However, this lower expression level resulted from the instability of the template and salt composition of the PCR mixture (Wang et al., 2018). Emerging strategies for improving the sfGFP synthesis yields, such as increasing the template length (Wu et al., 2007; Hong et al., 2014), or introducing specific chemical modifications (Hong et al., 2014; Sun et al., 2014), and DNA-binding proteins (Yim et al., 2020), did not achieve the desired result. To avoid laborious cloning and purification steps, we decided to directly use the PCR amplicons for measuring the translation initiation strength of the 5'-UTRs.

In theory, as the rate of 5'-UTR translation initiation increases, the protein expression level should increase accordingly. Here, we selected 93 5'-UTR variants from the random library and measured the sfGFP expression in a CFPS platform. To compare the influence of the 5'-UTRs on this different system, the experimental results were respectively normalized to the activity of the positive sequence. As shown (Figure 4B), the distributions of the relative translation activities between *in vivo* and *in vitro* systems were different. Overall, the 93 selected sequences were successfully expressed covering a 30-fold range of translation intensities, and most exhibited lower relative activity using the *in vitro* system. This result was consistent with previous reports that the PCR products were more susceptible to degradation by native exonucleases in the CFPS system (Michel-Reydellet et al., 2005; Seki et al., 2009). Moreover, the relative activity of the 5'-UTR random library showed a medium correlation in tuning the protein expression level between *in vivo* and *in vitro* systems (Pearson's $r = 0.72$, Figure 4C). Similarly, previous studies described a clear correlation between the relative strength of an RBS in *E. coli* cells and a CFPS system based on the cell extract (Chappell et al., 2013; Sun et al., 2014). More complex biological engineering, such as genetic circuit design (Sun et al., 2014) and the optimization of biosynthetic pathways (Liu et al., 2020) are also studied in cell-free systems. Additionally, based on high relative activity *in vivo*, the secondary structures of the five 5'-UTRs sequences that showed the highest and lowest activity *in vitro* were respectively identified and analyzed using NUPACK (Figure 4D), an online software for the prediction of nucleic acid structures. Compared to those with higher relative activity *in vitro*, the secondary structures of 5'-UTRs with lower relative activity were more complex (Osterman et al., 2013). As expected, 5'-UTRs with a weaker secondary structure performed the translation more efficiently, which could contribute to the interaction between the mRNA regions and the ribosome during the translation initiation process. This result implied that mRNAs containing 5'-UTR sequences with weak secondary structures tend to be translated more effectively *in vitro*, even though they are potentially more susceptible to degradation by exonucleases.

However, the related research indicated that there is a gap between the *in vivo* prediction and the real relative activity of *in vitro* systems, which were based on different bacterial chassis

strains (Wang et al., 2018; Zhang et al., 2021). The mentioned theoretical prediction results for 5'-UTRs were provided by the RBS calculator, which is a powerful platform enabling rational control of protein expression levels (Salis et al., 2009; Salis, 2011). Consistent with previous studies, we also observed a lower correlation between the *in vitro* and *in vivo* prediction results for the *E. coli* BL21 Star (DE3) strain (Pearson's $r = 0.31$, Supplementary Figure S5A). As further validation, the correlation of the actual translation initiation strength of 5'-UTRs in *E. coli* BL21 Star (DE3) with their corresponding predictions was low (Pearson's $r = 0.36$, Supplementary Figure S5B). This indicates that the influence of the 5'-UTRs on gene expression was not perfectly captured by the RBS calculator. It should be noted that the utilized promoter and plasmid were not suitable for the physical environment in the host cell system. The RBS calculator uses a theoretical thermodynamic model of Gibbs free energies of ribosome binding. While the optimal predicted sequence length before the start codon is 35 nucleotides in the RBS calculator, we only randomized 25 nucleotides. Overall, the observed differences between *in vitro* and *in vivo* systems may be due to disruption of the potential transcription and translation pathways and provides a general strategy for the design of regulatory elements in synthetic biology. In general, these results indicated that although cell-free systems can provide the advantage for rapid prototyping, further optimization of the cell-free systems needs to improve the directly reflective performance *in vivo*. Therefore, the development of an emerging platform for the precise prediction of 5'-UTRs should pay more attention to actual implementation in different expression systems (Na and Lee, 2010; Seo et al., 2013; Pandi et al., 2022). Looking forward, a novel strategy for characterizing biological systems based on the CFPS platform can provide a valuable tool for advancing the development of metabolic engineering and synthetic biology, similar to the comprehensive i^3 -screening pipeline (Kohyama et al., 2023).

4 Conclusion

In this study, we assessed the correlation of the translation initiation strength for diverse 5'-UTRs between *in vitro* and *in vivo* systems, which demonstrated differences in the distribution of relative activity and offered explanations of correlation analysis using the Pearson value. For the unbiased 5'-UTR library, a significant correlation of the absolute fluorescence between the *E. coli* K-12 and B strains was observed (Pearson's $r = 0.84$). For the biased 5'-UTR library, there were different degrees of correlation between two different types of chassis cells. In addition, an intermediate level of correlation of the relative activity of the 5'-UTR random library was demonstrated between the *in vivo* and *in vitro* systems (Pearson's $r = 0.72$). This result implies that there is some degree of deviation from 5'-UTR consistency between the living cell system and the cell-free system. More importantly, our analysis revealed that the lack of nucleotide C and complex secondary structure features in the 5'-UTR are beneficial to enhance the efficiency of protein expression in the different expression systems. Further studies can combine the flexibility and simplicity of cell-free systems to facilitate the manipulation of genetic elements for metabolic engineering and synthetic biology.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

JaL and PL contributed equally to this work. JaL and PL performed the main experiments and drafted the original manuscripts. QL conducted the statistical analysis. JnL collected the relevant subject literature. HQ conceived the experiments and supervised the study. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key R&D Program of China (Grant No. 2019YFA0904103).

References

- Arava, Y., Boas, F. E., Brown, P. O., and Herschlag, D. (2005). Dissecting eukaryotic translation and its control by ribosome density mapping. *Nucleic Acids Res.* 33, 2421–2432. doi:10.1093/nar/gki331
- Basu, I., Gorai, B., Chandran, T., Maiti, P. K., and Hussain, T. (2022). Selection of start codon during mRNA scanning in eukaryotic translation initiation. *Commun. Biol.* 5, 587. doi:10.1038/s42003-022-03534-2
- Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., and Kompala, D. S. (1990). Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol. Bioeng.* 35, 668–681. doi:10.1002/bit.260350704
- Carreiro, S., Ferreira, E. C., and Rocha, I. (2013). Metabolic responses to recombinant bioprocesses in *Escherichia coli*. *J. Biotechnol.* 164, 396–408. doi:10.1016/j.jbiotec.2012.08.026
- Chappell, J., Jensen, K., and Freemont, P. S. (2013). Validation of an entirely *in vitro* approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.* 41, 3471–3481. doi:10.1093/nar/gkt052
- Cuperus, J. T., Groves, B., Kuchina, A., Rosenberg, A. B., Jovic, N., Fields, S., et al. (2017). Deep learning of the regulatory grammar of yeast 5' untranslated regions from 500,000 random sequences. *Genome Res.* 27, 2015–2024. doi:10.1101/gr.224964.117
- Ding, W., Cheng, J., Guo, D., Mao, L., Li, J., Lu, L., et al. (2018). Engineering the 5' UTR-mediated regulation of protein abundance in yeast using nucleotide sequence activity relationships. *ACS Synth. Biol.* 7, 2709–2714. doi:10.1021/acssynbio.8b00127
- Duan, Y., Zhang, X., Zhai, W., Zhang, J., Zhang, X., Xu, G., et al. (2022). Deciphering the rules of ribosome binding site differentiation in context dependence. *ACS Synth. Biol.* 11, 2726–2740. doi:10.1021/acssynbio.2c00139
- Englaender, J. A., Jones, J. A., Cress, B. F., Kuhlman, T. E., Linhardt, R. J., and Koffas, M. A. G. (2017). Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth. Biol.* 6, 710–720. doi:10.1021/acssynbio.6b00350
- Evsratov, S. A., Osterman, I. A., Komarova, E. S., Pogorelskaya, A. M., Rubtsova, M. P., Zatsepin, T. S., et al. (2017). Application of sorting and next generation sequencing to study 5-UTR influence on translation efficiency in *Escherichia coli*. *Nucleic Acids Res.* 45, 3487–3502. doi:10.1093/nar/gkw1141
- Goodman, D. B., Church, G. M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479. doi:10.1126/science.1241934
- Han, L., Cui, W., Lin, Q., Chen, Q., Suo, F., Ma, K., et al. (2020). Efficient overproduction of active nitrile hydratase by coupling expression induction and enzyme maturation via programming a controllable cobalt-responsive gene circuit. *Front. Bioeng. Biotechnol.* 8, 193. doi:10.3389/fbioe.2020.00193
- Hodgman, C. E., and Jewett, M. C. (2012). Cell-free synthetic biology: Thinking outside the cell. *Metab. Eng.* 14, 261–269. doi:10.1016/j.ymben.2011.09.002
- Hong, S. H., Ntai, I., Haimovich, A. D., Kelleher, N. L., Isaacs, F. J., and Jewett, M. C. (2014). Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS Synth. Biol.* 3, 398–409. doi:10.1021/sb400140t
- Jeong, H., Barbe, V., Lee, C. H., Vallenet, D., Yu, D. S., Choi, S. H., et al. (2009). Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* 394, 644–652. doi:10.1016/j.jmb.2009.09.052
- Jiang, N., Ding, X., and Lu, Y. (2021). Development of a robust *Escherichia coli*-based cell-free protein synthesis application platform. *Biochem. Eng. J.* 165, 107830. doi:10.1016/j.bej.2020.107830
- Kightlinger, W., Lin, L., Rosztochy, M., Li, W., Delisa, M. P., Mrksich, M., et al. (2018). Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat. Chem. Biol.* 14, 627–635. doi:10.1038/s41589-018-0051-2
- Kim, D., Hong, J. S., Qiu, Y., Nagarajan, H., Seo, J. H., Cho, B. K., et al. (2012). Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genet.* 8, e1002867. doi:10.1371/journal.pgen.1002867
- Kohyama, S., Frohn, B. P., Babl, L., and Schwille, P. (2023). *Designing a protein with emergent function by combined in silico, in vitro and in vivo screening*. bioRxiv. doi:10.1101/2023.02.16.528840
- Komarova, A. V., Tchufistova, L. S., Dreyfus, M., and Boni, I. V. (2005). AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J. Bacteriol.* 187, 1344–1349. doi:10.1128/JB.187.4.1344-1349.2005
- Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255–258. doi:10.1126/science.1170160
- Kuhnert, P., Nicolet, J., and Frey, J. (1995). Rapid and accurate identification of *Escherichia coli* K-12 strains. *Appl. Environ. Microbiol.* 61, 4135–4139. doi:10.1128/aem.61.11.4135-4139.1995
- Kwon, Y. C., and Jewett, M. C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* 5, 8663. doi:10.1038/srep08663
- Li, G. W., Oh, E., and Weissman, J. S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484, 538–541. doi:10.1038/nature10965
- Liu, R., Zhang, Y., Zhai, G., Fu, S., Xia, Y., Hu, B., et al. (2020). A cell-free platform based on nisin biosynthesis for discovering novel lanthipeptides and guiding their overproduction *in vivo*. *Adv. Sci. (Weinh)* 7, 2001616. doi:10.1002/adv.202001616
- Lu, Y. (2017). Cell-free synthetic biology: Engineering in an open world. *Synth. Syst. Biotechnol.* 2, 23–27. doi:10.1016/j.synbio.2017.02.003
- Luli, G. W., and Strohl, W. R. (1990). Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Appl. Environ. Microbiol.* 56, 1004–1011. doi:10.1128/aem.56.4.1004-1011.1990
- Marisch, K., Bayer, K., Scharl, T., Mairhofer, J., Krempel, P. M., Hummel, K., et al. (2013). A comparative analysis of industrial *Escherichia coli* K-12 and B strains in high-glucose batch cultivations on process-, transcriptome- and proteome level. *PLoS One* 8, e70516. doi:10.1371/journal.pone.0070516

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2023.1201580/full#supplementary-material>

- Michel-Reydellet, N., Woodrow, K., and Swartz, J. (2005). Increasing PCR fragment stability and protein yields in a cell-free system with genetically modified *Escherichia coli* extracts. *J. Mol. Microbiol. Biotechnol.* 9, 26–34. doi:10.1159/000088143
- Na, D., and Lee, D. (2010). RBSDesigner: Software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics* 26, 2633–2634. doi:10.1093/bioinformatics/btq458
- Osterman, I. A., Evfratov, S. A., Sergiev, P. V., and Dontsova, O. A. (2013). Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.* 41, 474–486. doi:10.1093/nar/gks989
- Pandi, A., Diehl, C., Yazdizadeh Kharrazi, A., Scholz, S. A., Bobkova, E., Faure, L., et al. (2022). A versatile active learning workflow for optimization of genetic and metabolic networks. *Nat. Commun.* 13, 3876. doi:10.1038/s41467-022-31245-z
- Peng, X., Yu, K. Q., Deng, G. H., Jiang, Y. X., Wang, Y., Zhang, G. X., et al. (2013). Comparison of direct boiling method with commercial kits for extracting fecal microbiome DNA by Illumina sequencing of 16S rRNA tags. *J. Microbiol. Methods* 95, 455–462. doi:10.1016/j.mimet.2013.07.015
- Phue, J. N., Noronha, S. B., Hattacharyya, R., Wolfe, A. J., and Shiloach, J. (2005). Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: Differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and northern blot analyses. *Biotechnol. Bioeng.* 90, 805–820. doi:10.1002/bit.20478
- Phue, J. N., and Shiloach, J. (2004). Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in *E. coli* B (BL21) and *E. coli* K (JM109). *J. Biotechnol.* 109, 21–30. doi:10.1016/j.jbiotec.2003.10.038
- Pontrelli, S., Chiu, T. Y., Lan, E. I., Chen, F. Y., Chang, P., and Liao, J. C. (2018). *Escherichia coli* as a host for metabolic engineering. *Metab. Eng.* 50, 16–46. doi:10.1016/j.ymben.2018.04.008
- Posfai, G., Plunkett3rd, Feher, G., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046. doi:10.1126/science.1126439
- Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950. doi:10.1038/nbt.1568
- Salis, H. M. (2011). The ribosome binding site calculator. *Methods Enzymol.* 498, 19–42. doi:10.1016/B978-0-12-385120-8.00002-4
- Sample, P. J., Wang, B., Reid, D. W., Presnyak, V., McFadyen, I. J., Morris, D. R., et al. (2019). Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* 37, 803–809. doi:10.1038/s41587-019-0164-5
- Seidman, C. E., Struhl, K., Sheen, J., and Jessen, T. (2001). Introduction of plasmid DNA into cells. *Curr. Protoc. Mol. Biol. Chapter 1*, mb0108s37. Unit1 8. doi:10.1002/0471142727.mb0108s37
- Seki, E., Matsuda, N., and Kigawa, T. (2009). Multiple inhibitory factor removal from an *Escherichia coli* cell extract improves cell-free protein synthesis. *J. Biosci. Bioeng.* 108, 30–35. doi:10.1016/j.jbiosc.2009.02.011
- Seo, S. W., Yang, J. S., Kim, I., Yang, J., Min, B. E., Kim, S., et al. (2013). Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.* 15, 67–74. doi:10.1016/j.ymben.2012.10.006
- Shi, F., Luan, M., and Li, Y. (2018). Ribosomal binding site sequences and promoters for expressing glutamate decarboxylase and producing gamma-aminobutyrate in *Corynebacterium glutamicum*. *Amb. Express* 8, 61. doi:10.1186/s13568-018-0595-2
- Shiloach, J., Kaufman, J., Guillard, A. S., and Fass, R. (2000). Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by *Escherichia coli* BL21 (ADE3) and *Escherichia coli* JM109. *Biotechnol. Bioeng.* 49, 421–428. doi:10.1002/(sici)1097-0290(19960220)49:4<421::aid-bit9>3.0.co;2-r
- Shiloach, J., Reshamwala, S., Noronha, S. B., and Negrete, A. (2010). Analyzing metabolic variations in different bacterial strains, historical perspectives and current trends—example *E. coli*. *Curr. Opin. Biotechnol.* 21, 21–26. doi:10.1016/j.copbio.2010.01.001
- Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755. doi:10.1038/90802
- Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U. S. A.* 71, 1342–1346. doi:10.1073/pnas.71.4.1342
- Son, Y. J., Phue, J. N., Trinh, L. B., Lee, S. J., and Shiloach, J. (2011). The role of Cra in regulating acetate excretion and osmotic tolerance in *E. coli* K-12 and *E. coli* B at high density growth. *Microb. Cell Fact.* 10, 52. doi:10.1186/1475-2859-10-52
- Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2014). Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* 3, 387–397. doi:10.1021/sb400131a
- Verma, M., Choi, J., Cottrell, K. A., Lavagnino, Z., Thomas, E. N., Pavlovic-Djuranovic, S., et al. (2019). A short translational ramp determines the efficiency of protein synthesis. *Nat. Commun.* 10, 5774. doi:10.1038/s41467-019-13810-1
- Wang, H., Li, J., and Jewett, M. C. (2018). Development of a *Pseudomonas putida* cell-free protein synthesis platform for rapid screening of gene regulatory elements. *Synth. Biol. (Oxf)* 3, ysy003. doi:10.1093/synbio/ysy003
- Wu, P. S., Ozawa, K., Lim, S. P., Vasudevan, S. G., Dixon, N. E., and Otting, G. (2007). Cell-free transcription/translation from PCR-amplified DNA for high-throughput NMR studies. *Angew. Chem. Int. Ed. Engl.* 46, 3356–3358. doi:10.1002/anie.200605237
- Wu, Y., Tang, M., Wang, Z., Yang, Y., Li, Z., Liang, S., et al. (2022). Efficient *in vitro* full-sense-codons protein synthesis. *Adv. Biol. (Weinh)* 6, e2200023. doi:10.1002/adbi.202200023
- Yim, S. S., Johns, N. L., Noireaux, V., and Wang, H. H. (2020). Protecting linear DNA templates in cell-free expression systems from diverse bacteria. *ACS Synth. Biol.* 9, 2851–2855. doi:10.1021/acssynbio.0c00277
- Yoon, S. H., Han, M. J., Jeong, H., Lee, C. H., Xia, X. X., Lee, D. H., et al. (2012). Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biol.* 13, R37. doi:10.1186/gb-2012-13-5-r37
- Yu, B. J., Sung, B. H., Koob, M. D., Lee, C. H., Lee, J. H., Lee, W. S., et al. (2002). Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. *Nat. Biotechnol.* 20, 1018–1023. doi:10.1038/nbt740
- Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., et al. (2011). NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* 32, 170–173. doi:10.1002/jcc.21596
- Zhang, L., Lin, X., Wang, T., Guo, W., and Lu, Y. (2021). Development and comparison of cell-free protein synthesis systems derived from typical bacterial chassis. *Bioresour. Bioprocess* 8, 58. doi:10.1186/s40643-021-00413-2